July 09, 2007

From
Shivkumar Kalyanaraman
Professor, Dept. of Electrical, Computer and Systems Engineering (ECSE),
Rensselaer Polytechnic Institute (RPI), Troy NY 12180
Phone: (518) 276 8979 ; Email: shivkuma@ecse.rpi.edu
and
Murat Yuksel
Assistant Professor, Dept. of Computer Science & Engineering,
University of Nevada, Reno,
1664 N. Virginia St., Reno, NV 89557-0042
Phone: +1 (775) 327 2246; Email: yuksem@cse.unr.edu

To
Marlene Dortch
Federal Communications Commission Secretary
445 12th Street, SW
Washington DC 20554

Re: WC Docket No. 07-52

Dear Ms. Dortch,

We write in response to the FCC's Notice of Inquiry on broadband industry practices. We wish to bring to your attention our recent study, *"Value of Class-of-Service Support in IP Backbones."* The manuscript has been attached with this letter. A shorter and earlier version of this manuscript has been published in IEEE International Workshop on Quality of Service (IwQoS) 2007, a peer-reviewed workshop. This study quantifies the relative extra capacity needed between simple two-class networks and classless networks holding the service level requirements equivalent to what higher priority service experiences. Our study makes clear that there are substantial additional extra capacity required to operate networks in which all traffic is treated alike, and carrying traffic that needs to still be assured performance as specified in service level agreements (SLAs).

As the Internet becomes more crowded with high-bandwidth applications and content, the issue of "network neutrality" involves both economic and technical aspects. One aspect of the debate involves whether application traffic that requires performance assurances (e.g., VoIP) could be serviced differently, or what the impact would be if all traffic were to be treated in an undifferentiated manner. Thus our study attempts to answer one basic question: "If we want to meet the needs of applications that require service level assurances, how much more capacity do I need?" We compared the current "best-effort" approach with a tiered model that separates information into two simple classes — one for most types of information and another for applications requiring service level assurance for high-bandwidth content like video games, telemedicine, and Voice over Internet Protocol (VoIP). We show that an over-provisioned network will require significant excess capacity compared to a simple two-class network with the same traffic loads, but protects the service levels of rich media applications. While the amount of required excess capacity is contingent on several factors explored in the study, it can be quite significant even for moderate network utilization, and even for small fractions of rich media traffic compared to regular internet traffic. In summary, over-provisioning requires substantial excess capacity, given the assumptions in our study.

We thank the FCC for its careful consideration of broadband policy issues. The opinions expressed above are our own and do not reflect the official position of our employers.

Sincerely,

K. Shivkumar

Shivkumar Kalyanaraman
Professor, Dept of ECSE,
Rensselaer Polytechnic Institute

Murat Yuksel
Assistant Professor, Dept of CSE
University of Nevada, Reno

Value of Supporting Class-of-Service in IP Backbones

Murat Yuksel, University of Nevada - Reno, yuksem@cse.unr.edu

K. K. Ramakrishnan, AT&T Labs Research, kkrama@research.att.com

Shivkumar Kalyanaraman, Rensselaer Polytechnic Institute, shivkuma@ecse.rpi.edu

Joseph D. Houle, Rita Sadhvani, AT&T, {jdhoule,rsadhvani}@att.com

***Abstract*—The desire or ability of an ISP to provide differentiated service is a current hotly debated topic. In this paper, we quantify the value of having differentiated service (i.e., class-of-service (CoS)) support in an IP backbone. We compare the capacity requirements of a Diffserv environment providing service for applications that require delay or loss assurances in comparison to a network that provides classless (i.e., best-effort) service and still has to meet the same performance assurances. Our modeling framework first develops a link model that quantifies the Required Extra Capacity (REC) in order for a classless link to provide the same level of performance as experienced by premium class traffic passing through a fixed capacity CoS link. We develop the REC calculations for the cases when average delay or the average loss probability is the target performance goal with Poisson or Markov Modulated Poisson Process (MMPP) input traffic. Our primary contribution is in quantifying the value of the CoS support in a network setting.**

I. Introduction

Currently there is a wide ranging debate about the issue of "network neutrality" which involves both economic and technical aspects [1], [2]. One key technical aspect of the network neutrality debate is whether best effort application traffic should be carried along with other (so-called "premium") traffic for which SLA commitments have been made (or are expected, either explicitly or implicitly). An assertion often made in this context is that over-provisioning is an economically viable strategy due to the declining cost of capacity, instead of incurring the complexity and operational costs of running a differentiated services network. Our study focuses on this specific question within the larger debate. We compare a classless network which is over-provisioned against an engineered network using per-class queuing to offer Class-of-Service (CoS) and meet user expectations and SLAs.

There is a vast body of network QoS literature studying different queueing, scheduling and buffer management mechanisms to allocate finite capacity and delay (given an average utilization) amongst flows at a statistically multiplexed resource [3]. Recent work by Ciucu et al [4] proposes a provisioning strategy based upon statistical service curve characterization and argues that scheduling has little value added above such provisioning. In our work, a key difference is that we do not have admission control or shaping/policing of input traffic, but since the network must still honor premium SLAs, simple CoS scheduling is valuable. The flow-aware networking approach [5] suggests the use of implicit differentiation by using per-flow queuing and per-flow admission control. In contrast, our work focusses on comparing a simple 2-class vs. 1-class model at the aggregate level without admission control. An analysis similar to ours was done by Sahu et al. [6] in comparing loss performance of forwarding behaviors (i.e., discard eligibility vs. priority) of the diff-serv architecture. Instead of services specific to the diff-serv architecture, our work compares the classless service to the class-of-service in general. We also provide the quantitative comparison at the edge-to-edge (g2g) level with full consideration of network-specific issues such as the topology and the traffic matrix. Huang and Guerin [7] have examined the benefit of over-provisioning to overcome traffic uncertainty and to accommodate scaling up of the network. In contrast, we examine the relative benefit of CoS support rather than over-provisioning the network. QoS research has also recognized the need to *simplify and de-couple* QoS building blocks to promote implementation and inter-network deployment. The IETF Int-serv [8] and Diff-serv [9] models simplify the architecture for supporting Class-of-Service (CoS) in the core IP network.

The rest of the paper is organized as follows: We first describe our modeling framework in Section II. Then, to quantify REC, we detail link models for Poisson and MMPP traffic cases. In Section IV, we extend the link models to a network model using sample ISP topologies. We present the REC for these ISP topologies required to meet the performance needed by some existing applications. We then conclude.

II. Model Framework

The goal of our modeling study is to provide a framework where capacity provisioned for a classless service can be easily compared to that for a CoS based network for various ISP topologies. We start by considering two traffic classes on a single CoS link: *premium traffic class* and *best-effort traffic class*. We, then, set a performance goal of delay or loss for the premium traffic on the CoS link, and then seek to find the required extra capacity (REC) for a classless link (which treats both traffic classes equally) to achieve the same performance goals for both the traffic classes. Figure 1 illustrates the comparison of the two options for the service provided at the link. Let the aggregate traffic rate be λ_D to be served by a CoS link with a capacity of μ_D. Also let a fraction of this aggregate traffic be premium class traffic with a rate of $\lambda_{Prem} = g\lambda_D$ with the rest of the aggregate traffic being best-effort (BE) class with a rate of $\lambda_{BE} = (1-g)\lambda_D$. For the premium class

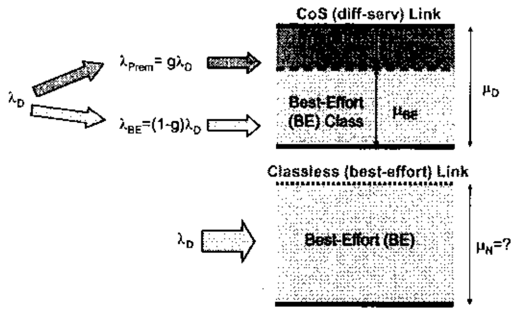


Fig. 1. Link-level comparison of two service types: Classless vs. CoS.

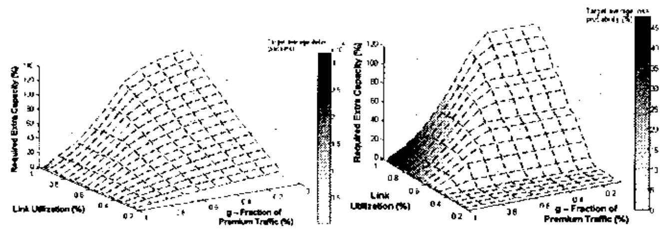

(a) $a = 0.5$, $r = 4$ (b) $a = 0.5$, $r = 4$, $K = 60$

Fig. 2. Link model results: (a) MMPP/M/1: REC to achieve target delay, (b) MMPP/M/1/K: REC to achieve loss rate. The surface darkness shows the target.

traffic, we define a performance goal ζ, e.g., average delay t_{target} or average packet loss probability p_{target}.

Given the parameters as illustrated in Figure 1, the model aims to formulate the necessary classless link capacity μ_N to achieve the same performance goal ζ for the aggregate traffic λ_D. From this, REC for the CoS link can be calculated in terms of *rate* as $\mu_N - \mu_D$ (or as a *percentage* $100(\mu_N/\mu_D - 1)$). With this model, one can use average delay t_{target} or average loss probability p_{target} as the performance goal. In the case of loss probability, an additional parameter is the buffer size K. Since *non-preemptive priority queuing* is a simple, analytically tractable packet scheduling policy for CoS support, we base our analysis on it. We develop our link model based on two different traffic models: Poisson traffic and a Markov-Modulated Poisson Process (MMPP) [10].

III. Link Model

A. Poisson Traffic

1) Achieving a Delay Target: M/M/1 Model: The first scenario we model is the case when traffic is assumed to be Poisson and target performance goal is determined in terms of delay, i.e., t_{target}. Let μ_N be the required capacity for the classless link to be able to match the premium class performance in CoS. The delay achieved by the classless service for the aggregate traffic will be:

$$t_{achieved} = \frac{1}{\mu_N - \lambda_D} \tag{1}$$

When the achieved delay $t_{achieved}$ is equal to the target delay (i.e., $t_{target} = t_{achieved}$), the classless link capacity μ_N equals to the *minimum* required to satisfy the performance goal:

$$\mu_N = \frac{1}{t_{target}} + \lambda_D \tag{2}$$

Equation 2 shows that the REC depends on the rigor of the performance goal $1/t_{target}$ and the aggregate traffic rate λ_D of the CoS link. However, not all values of t_{target} might be achievable for the premium class traffic at the CoS link. The average delay that the premium class experiences at the CoS link is dependent on three factors: (i) the aggregate traffic rate λ_D, (ii) the fraction g of the premium class in the aggregate traffic, and (iii) the CoS link capacity μ_D.

2) Achieving a Loss Target: M/M/1/K Model: We now look at the case when the performance target is determined in terms of average packet loss probability, p_{target}, where we consider the buffer size K. We assume that CoS link provides an equal buffer of K packets to both traffic classes, and that the classless link uses both the buffers (i.e., total of $2K$ packets) for the aggregate traffic. As the buffer is bounded with $2K$ packets, the average loss probability achieved by the classless service for the aggregate traffic can be approximated by the tail probability of the queue:

$$p_{achieved} = \left(\frac{\lambda_D}{\mu_N}\right)^{2K} \tag{3}$$

When the achieved loss probability $p_{achieved}$ is equal to the target performance p_{target} (i.e., $p_{target} = p_{achieved}$), the classless link capacity equals the *minimum* required to satisfy the performance goal:

$$\mu_N = \lambda_D \frac{1}{\sqrt[2K]{p_{target}}} \tag{4}$$

Similar to the previous case, Equation 4 shows that the REC depends on the rigor of the performance goal $1/p_{target}$ and the aggregate traffic rate λ_D of the CoS link. It now also depends on the available buffer size of the classless link, i.e., $2K$.

B. MMPP Traffic

To approach a more realistic traffic model, we use a generic Markov-Modulated Poisson Process (MMPP) [10] with two states ($i = 1, 2$) each corresponding to a particular sending rate λ_i of a Poisson process, with a target average sending rate of λ_t. As the ratio λ_2/λ_1 gets higher, the generated traffic becomes more bursty. Let the sending rate of the first state be a fraction $0 < a < 1$ of the average rate (i.e. $\lambda_1 = a\lambda_t$) and let r be the ratio of the traffic rates of the two MMPP states (i.e., $r = \lambda_2/\lambda_1$). To generate traffic with an average rate λ_t, we set the two traffic rates as

$$\lambda_1 = a\lambda_t, \pi_1 = (ar-1)/(ar-a)$$
$$\lambda_2 = ar\lambda_t, \pi_2 = (1-a)/(ar-a)$$

where $0 < a < 1$, $r > 1/a$, and π_1 and π_2 are the state probabilities.

We used ns-2 simulations of a classless and CoS link to obtain an accurate link model for the MMPP traffic case which is used in the network scenarios. To simulate the CoS link, we use priority queuing for flows from two classes passing through the link. For the classless link simulation, we used a FIFO queue for a single flow with a rate equal to the aggregate of the two flows of the CoS case, reflecting the analytical link models. In order to find the REC values by simulation, we match the performance experienced by the premium class flow in the CoS link with the one experienced by the single flow over the classless link, within 1% error (6 repetitions of each case). We then generate REC values based on these simulation of the link model for MMPP/M/1 and MMPP/M/1/K cases.

As we see in Figure 2, the REC grows as the link utilization becomes higher, but more so when the fraction of premium traffic (g) is smaller. This is important as we expect that the network will likely see a slowly increasing amount of premium class traffic. But even at $g = 0.5$, the REC can be quite significant (e.g., 50%) at high utilizations (e.g., 0.8).

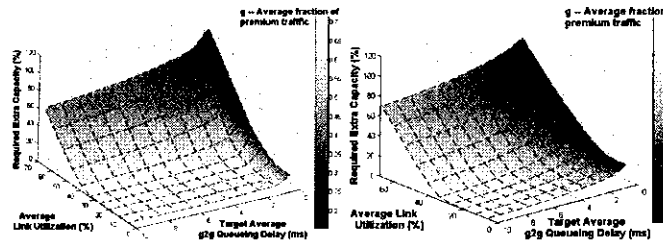
IV. Network Model

The final step involves generalizing the single link model into a network model. We focus on developing our network model for a typical ISP's backbone network. Briefly, we first calculate a routing matrix R for the ISP network from the link weight information. Given that a realistic traffic matrix T is available, we then calculate the traffic load pertaining to individual links by performing the product of T and R which shows the distribution of traffic loads on individual links. For each of these link traffic loads, the link model described earlier will apply. The link-load distribution will thus lead to a distribution of REC over links for the network.

The goal of the network model is to determine the *additional percentage capacity needed for a classless network over a CoS network* on an edge-to-edge (g2g) basis. The network model takes the following steps to calculate the network REC:

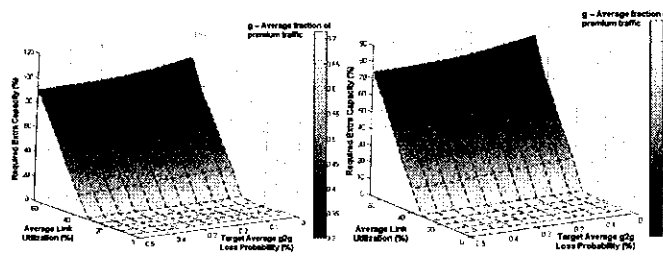
- *Step 1:* Construct the *routing matrix* $R_{F\times L}$ based on shortest path first (Diskstra's) algorithm.
- *Step 2:* Form the *traffic vector* $A_{F\times 1}$.
- *Step 3:* Calculate the *traffic load on each link* by performing the matrix operation $Q = R^T\lambda$, where $Q_{L\times 1}$ is the link load vector (in Mb/s).
- *Step 4:* Check and fix the *feasibility* of the traffic load and routing.
- *Step 5:* Calculate the *per-link* REC by using Q_i as the traffic rate λ_D for the ith link and the performance goal t_{target} or p_{target} for that link i.
- *Step 6:* Calculate the *network* REC by averaging the per-link RECs from Step 5.

To obtain realistic topology information for ISPs, we used the Rocketfuel [11] data repository which provides router-level topology data. In order to assign estimated capacity values for individual links, we used a technique based on a Breadth-First Search (BFS) algorithm. We first select the maximum-degree router in the topology as the source node for BFS to start. Then we assign higher capacities to the links close to this max-degree router. We use a gravity model [12] to construct a feasible traffic matrix composed of edge-to-edge (g2g) flows. We select the routers with lower degree and higher distance



(b) Abovenet (c) Sprintlink

Fig. 3. MMPP/M/1 network model: The surface darkness shows the g2g target queueing delay. MMPP's burstiness is defined by $a = 0.5$ and $r = 4$.



(b) Abovenet (c) Sprintlink

Fig. 4. MMPP/M/1/K network model: The surface darkness shows the g2g target loss probability. The buffer size is $K = 60$ packets and the MMPP's burstiness is defined by $a = 0.5$ and $r = 4$.

to the max-degree node as edge routers originating traffic into the network. We also ensure that there exists at least one edge router for each PoP in the topology. In the gravity model, we used CIESIN [13] dataset to calculate the city populations.

To evolve from the link model of Section III to the network model, we split the g2g performance goals on individual links of the g2g path. In order to split g2g delay target t_{target} on individual links, we simply divide the delay requirement equally on each link of the path assuming that t_{target} is only the delay in queueing and insertion into the links. After *equally* splitting the g2g delay on individual links for all g2g flows, we collect the tightest (i.e., minimum) delay requirement on each individual link among the delay requirements imposed by each g2g flow traversing the link. Similarly, given a g2g loss probability target p_{target}, we assign the loss probability requirement on each link of the path as follows. Specifically, for a path with l links, we *equally* assign the survival probability to each link as the lth root of the overall path's survival probability $1 - p_{target}$.

V. Network Model Results

Based on the network model from the previous section, we present *network REC* for selected Rocketfuel topologies for MMPP-based traffic models. In order to generate the network REC results, we use link model REC results for a given utilization and performance target. We perform a lookup from the simulation results of the link models (i.e., MMPP/M/1 and MMPP/M/1/K) and a linear interpolation on the link model REC values using the available datapoints to obtain the appropriate link REC. We use $a = 0.5$ and $r = 4$ as parameters for the MMPP traffic model. Real IP traffic is considered to be more bursty than what these values represent [14]. Also, when loss probability is the performance goal, we use a buffer size of $K = 60$ packets. For delay as the primary performance metric and since the g2g delay can be reduced by smaller buffer sizes [15], we chose a relatively small buffer of 120 packets (i.e., $K = 60$ packets, corresponding to a buffer time 100ms and 1ms for 10Mb/s and 1Gb/s links respectively with 1KB packets). Also, we use 0.1-10ms and 0.01-0.5% as the g2g queuing delay (excluding propagation) and loss probability targets respectively, as these are the required performance ranges for current and potential network applications.

Figure 3 shows the behavior of average REC across all links for two different network topologies, when the target edge-to-edge delay is the criterion, for the MMPP traffic case. Note that the fourth dimension, reflected in the shading and the vertical bar on the right, represents g, the proportion of premium traffic. As the average link utilization goes up, the REC goes up, especially when the target average delay is smaller. As we see, when g is small, REC is higher, because the extra capacity needed for the classless service has to be higher to ensure that the arrivals for the premium class are served as quickly as the CoS case would, with non-preemptive priority scheduling. Although the exact amount of REC changes with each topology, the REC needed on each link is similar across the topologies (in the range of 50-100% at average link utilizations of 80%).

Figure 4 shows how the average REC changes for the two topologies for the edge-to-edge packet loss criterion. We see across the network topologies, the REC increases as the utilization increases beyond a threshold (below which the buffering enables the classless service to avoid losses), and increases as the proportion g becomes smaller (darker shade). However, the increase in the REC is not as rapid when the target loss probability reduces from 0.5% down to 0.1%, which again reflects the role of buffering at each of the links. It is important to note however, that for average link utilizations of 80%, the average REC can be up to 100% for the case examined in the figure. It is important to note that under failure situations link utilizations can easily get well over 80% carrying protection traffic, even in a well-engineered and provisioned network.

VI. Summary

There has been a large body of work on supporting Quality of Service in the network, and quantifying the benefits in terms of reducing both the magnitude and variability of delay and loss experienced in the network. On the other hand, there has also been considerable debate on the benefits of having a simple classless (i.e., un-differentiated, single-class) service. However, the quantification of the amount of extra capacity required of such a classless network to support traffic that requires delay and loss service-level assurances has not been explored in the past. In this paper, we have quantified the required extra capacity (REC) for a classless network to meet the same delay and loss assurances that would be provided by a relatively simple two-class diff-serv based network.

We first built an analytical framework to understand the nature of REC for a link, which we feel is novel and interesting in its simplicity. Using Poisson traffic and exponential service times (M/M/1), we demonstrated the nature of the ratio in simple analytic terms, while recognizing the over-simplification of using such a traffic model. We then used a more realistic two state MMPP traffic arrival process to quantify REC for a link. We observed that REC grows with utilization, and is of particular concern when the proportion of premium class traffic requiring delay or loss assurances is small. The REC grows as the traffic becomes more bursty.

We validated our analytical results with simulation, and then used the results from the link level simulation experiments as the basis for showing the behavior of REC with more complex and realistic network models for an IP backbone. We used IP backbone topologies from Rocketfuel along with a careful and rigorous procedure for synthetically generating traffic matrices based on relative user populations while ensuring link capacities are sufficient to support the traffic. We observe that the REC increases with the average utilization of the links in the network and as the relative proportion g of the premium traffic reduces. Moreover, REC grows rapidly as the acceptable delay and packet loss targets become tighter (smaller). So for example, with conservative assumptions on the burstiness of the traffic (2-state MMPP parameters), REC approaches 60% even at reasonable average link utilizations of 60%, for a relatively small proportion (e.g., 20%) of premium class traffic.

References

[1] "Network Neutrality," http://en.wikipedia.org/wiki/Network_neutrality.

[2] J. Crowcroft, "Net Neutrality: The Technical Side of the Debate: A White Paper," *ACM Computer Communications Review (CCR)*, vol. 37, no. 1, pp. 49–55, January 2007.

[3] G. Cortese, R. Fiutem, P. Cremonese, S. D'Antonio, M. Esposito, S. P. Romano, and A. Diaconescu, "CADENUS: Creation and deployment of end-user services in premium IP-networks," *IEEE Communications Magazine*, vol. 41, no. 1, pp. 54–60, Jan 2003.

[4] F. Ciucu, A. Burchard, and J. Liebeherr, "A network service curve approach for the stochastic analysis of networks," in *Proc. of ACM SIGMETRICS*, New York, NY, USA, 2005, pp. 279–290.

[5] S. Oueslati and J. Roberts, "A new direction for quality of service: Flow aware networking," in *Proc. of NGI*, 2005, pp. 1389–1399.

[6] S. Sahu, D. Towsley, and J. Kurose, "A quantitative study of differentiated services for the Internet," in *Proc. of IEEE GLOBECOM*, 1999.

[7] Y. Huang and R. Guerin, "Does Over-Provisioning Become More or Less Efficient as Networks Grow Larger?," in *Proc. of ICNP*, 2005.

[8] R. Braden and et al, "Integrated services in the Internet architecture: An overview," *IETF RFC 1633*, Jun 1994.

[9] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," *IETF RFC 2475*, Dec 1998.

[10] W. Fischer and K. Meiher-Hellstern, "The Markov-Modulated Poisson Process (MMPP) cookbook," *Performance Evaluation*, vol. 18, pp. 149–171, 1992.

[11] N. Spring, R. Mahajan, and D. Wetherall, "Measuring ISP Topologies with Rocketfuel," in *Proc. of ACM SIGCOMM*, 2002.

[12] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic Matrix Estimation: Existing Techniques and New Directions," in *Proc. of ACM SIGCOMM*, 2002.

[13] "The Center for International Earth Science Information Network (CIESIN)," http://www.ciesin.columbia.edu.

[14] H. Jiang and C. Dovrolis, "Why is the Internet Traffic Bursty in Short Time Scales?," in *Proc. of ACM SIGMETRICS*, 2005.

[15] G. Appenzeller, I. Keslassy, and N. McKeown, "Sizing router buffers," in *Proc. of ACM SIGCOMM*, 2004.